Sentiment Analysis

On

Amazon Food Product Reviews

Jiali Han

han.jial@northeastern.edu

Content

01 Introduction

Background
Overview
Objectives

02 Related Works

Previous studies
Gaps in research

03 Methodology

Data collection
Feature extraction
Classification algorithms
DistilBERT fine-tuning
Evaluation Metrics

04 Experiments

Environment setup
Process flow

05 Results

Model performance
Discussion analysis

06 Conclusion

Ethics
Limications
Future work

Background

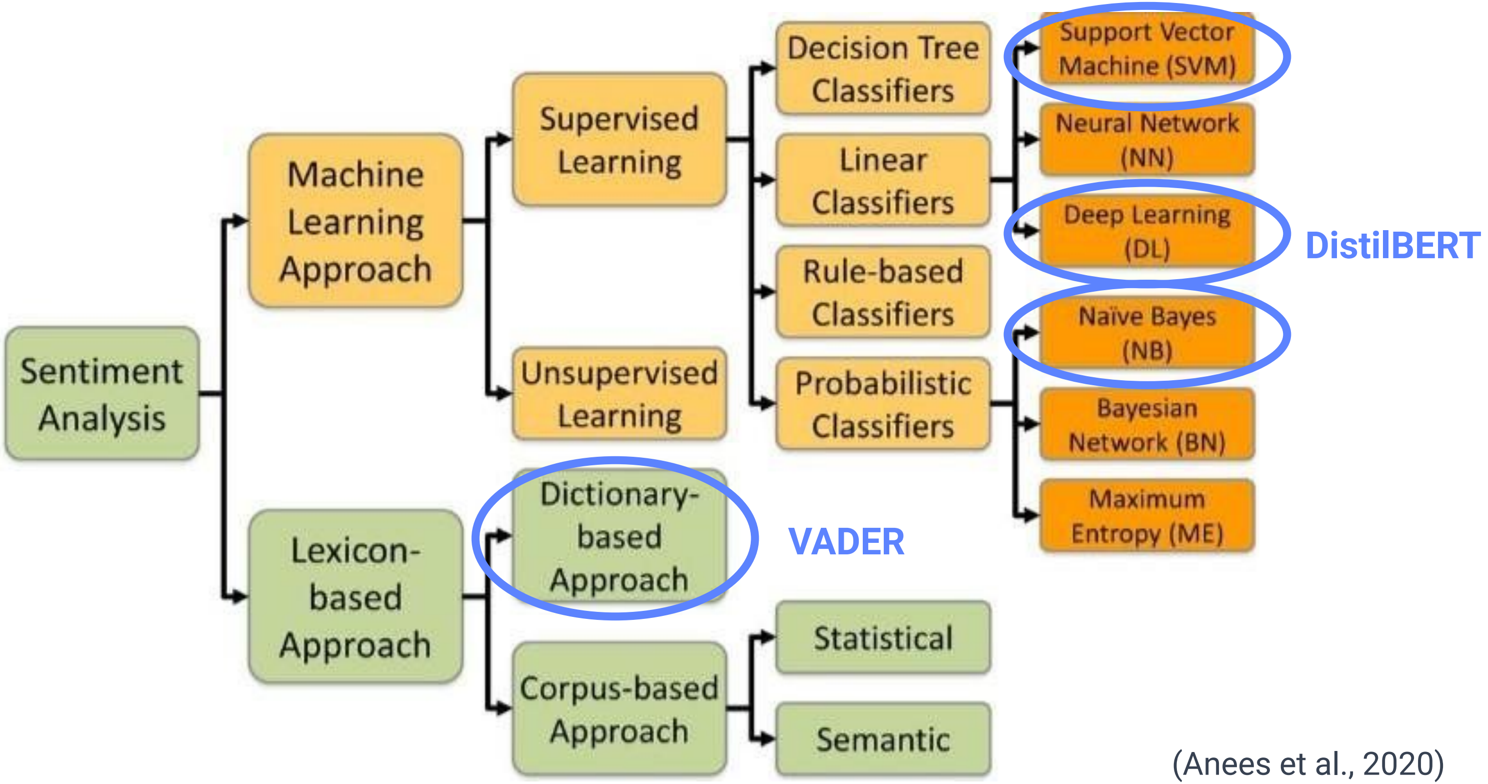
- **Sentiment Analysis in NLP**
- Sentiment Analysis in E-commerce
- Challenges and Evolution

(Anees et al., 2020)

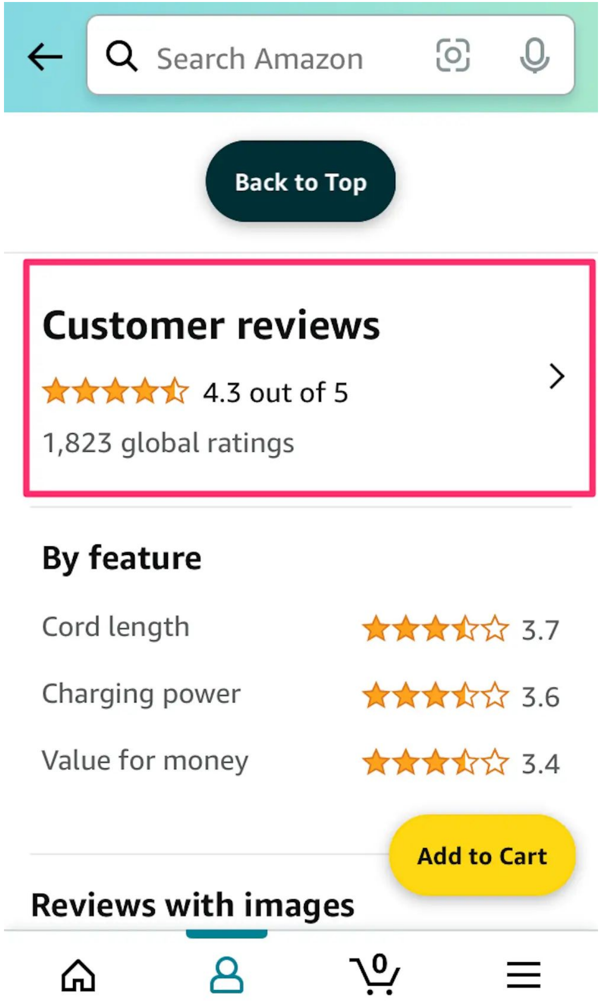
Background

- **Sentiment Analysis in NLP**

Logistic Regression (LR)

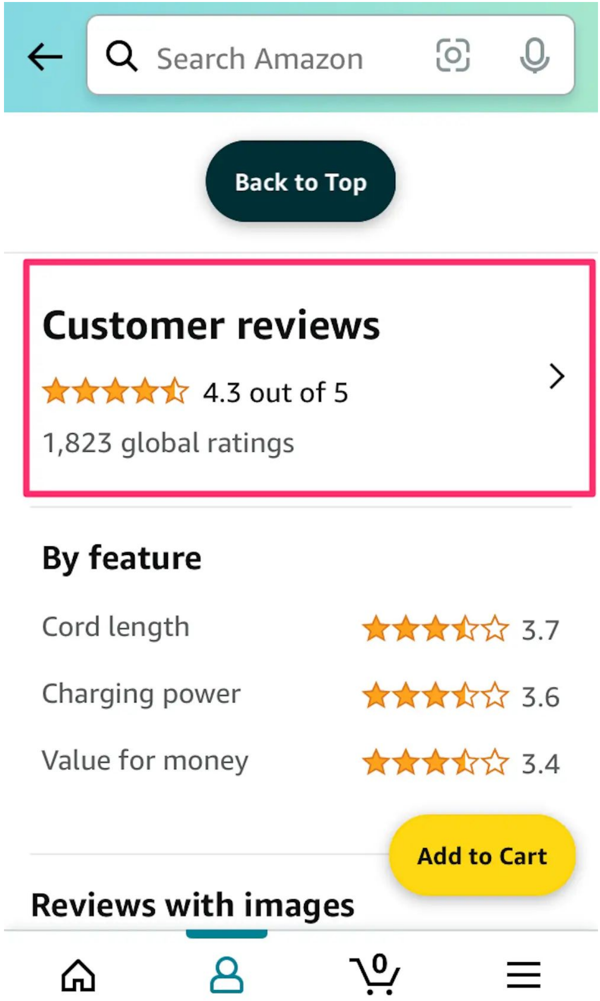


(Anees et al., 2020)



Background

- Sentiment Analysis in NLP
- **Sentiment Analysis in E-commerce**
- Challenges and Evolution



Background

- Sentiment Analysis in NLP
- Sentiment Analysis in E-commerce
- **Challenges and Evolution**

Classification of Amazon food product review

Goya Lady Fingers 7.0 OZ › Customer reviews

Customer reviews

★★★★☆ 4.4 out of 5

321 global ratings

Goya Lady Fingers 7.0 OZ

by Goya

5 star	73%
4 star	12%
3 star	7%
2 star	2%
1 star	6%

Write a review

How customer reviews and ratings work

Top positive review

Positive reviews ›

Gambit

★★★★★ **Perfect**

Reviewed in the United States on September 21, 2023

Exactly what you want from lady fingers. Crispy, tasty, dry.

Top critical review

Critical reviews ›

Daniel

★★★☆☆ **Expensive**

Reviewed in the United States on January 1, 2022

Way to expensive, walmart has the same package for $1.24.
I understand it may be the commodity, that you don't have to go to the store and search for them, but at the same time mine arrived all in pieces and the ones from walmart were all good .
For me it was a waste of money, but if you don't have an option I guess is ok to try your luck.

6 people found this helpful

Scope

- Sentiment Analysis of Amazon food product reviews
- English-language review texts
- Excludes non-textual and non-English reviews

Objectives

- Model Evaluation
- Model Innovation
- Comparative Analysis
- Results and Impact
- Future Direction and Applications

Related Works

2009 & 2012

Movie Reviews

Feature selection methods and ML classifiers like SVM, Naive Bayes, etc.

2011 &2015

Twitter

Advanced methods like POS-enhanced and unsupervised model with sentiment lexicons

2015 & 2018

Product reviews

Supervised models like perceptron Naives Bayes, SVM, Logistic Regression, Random Forest

2019 & 2020

Evolutions

Deep-learning neural networks like RNNs, CNN, and BiGRU

Despite advancements, there remains a **research gap** in accurately capturing the nuanced sentiment expressions specific to Amazon product reviews.

Content

01 Introduction
Background
Overview
Objectives

02 Related Works
Previous studies
Gaps in research

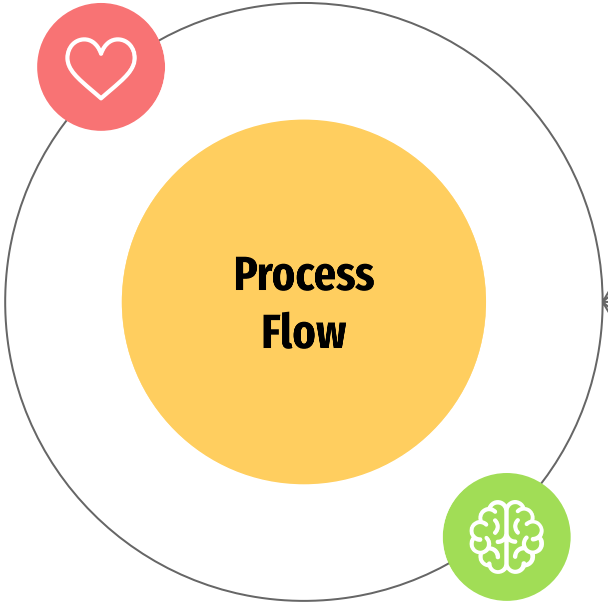
03 Methodology
Data collection
Feature extraction
Classification algorithms
DistilBERT fine-tuning
Evaluation Metrics

04 Experiments
Environment setup
Process flow

05 Results
Model performance
Discussion analysis

06 Conclusion
Ethics
Limications
Future work

Methodology



01 Data Collection and preprocessing

02 Feature Extraction Methods

03 Classification Tool / Algorithms

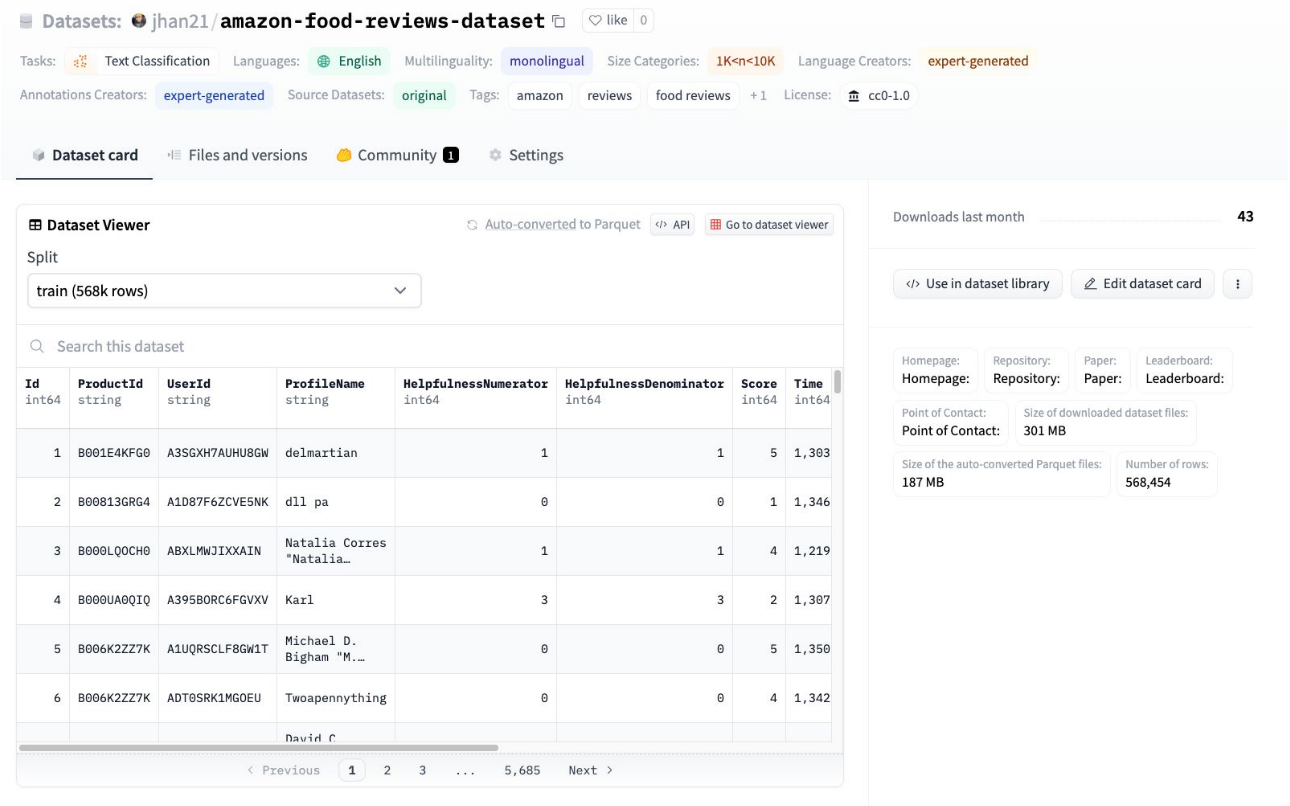
04 Fine-tuned DistilBERT Model

05 Evaluation Metrics

Data Collection and Preprocessing

- Data Compilation
- Preprocessing and Cleaning
- Class Transformation for Sentiment Analysis
- Word Cloud Analysis and Sentiment Complexity
- Dataset Partitioning and Resampling Techniques

Dataset



Datasets: jhan21/**amazon-food-reviews-dataset** like 0

Tasks: Text Classification Languages: English Multilinguality: monolingual Size Categories: 1K<n<10K Language Creators: expert-generated

Annotations Creators: expert-generated Source Datasets: original Tags: amazon reviews food reviews +1 License: cc0-1.0

Dataset card Files and versions Community 1 Settings

Dataset Viewer Auto-converted to Parquet API Go to dataset viewer

Split

train (568k rows)

Search this dataset

Id int64	ProductId string	UserId string	ProfileName string	HelpfulnessNumerator int64	HelpfulnessDenominator int64	Score int64	Time int64
1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1,303
2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1,346
3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia…	1	1	4	1,219
4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1,307
5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M.…	0	0	5	1,350
6	B006K2ZZ7K	ADT0SRK1MGOEU	Twoapennything	0	0	4	1,342
			David C				

< Previous 1 2 3 ... 5,685 Next >

Downloads last month 43

Use in dataset library Edit dataset card

Homepage: Homepage:
Repository: Repository:
Paper: Paper:
Leaderboard: Leaderboard:
Point of Contact: Point of Contact:
Size of downloaded dataset files: 301 MB
Size of the auto-converted Parquet files: 187 MB
Number of rows: 568,454

Dataset Card for "Amazon Food Reviews"

Hugging Face

Dataset Card

https://huggingface.co/datasets/jhan21/amazon-food-reviews-dataset

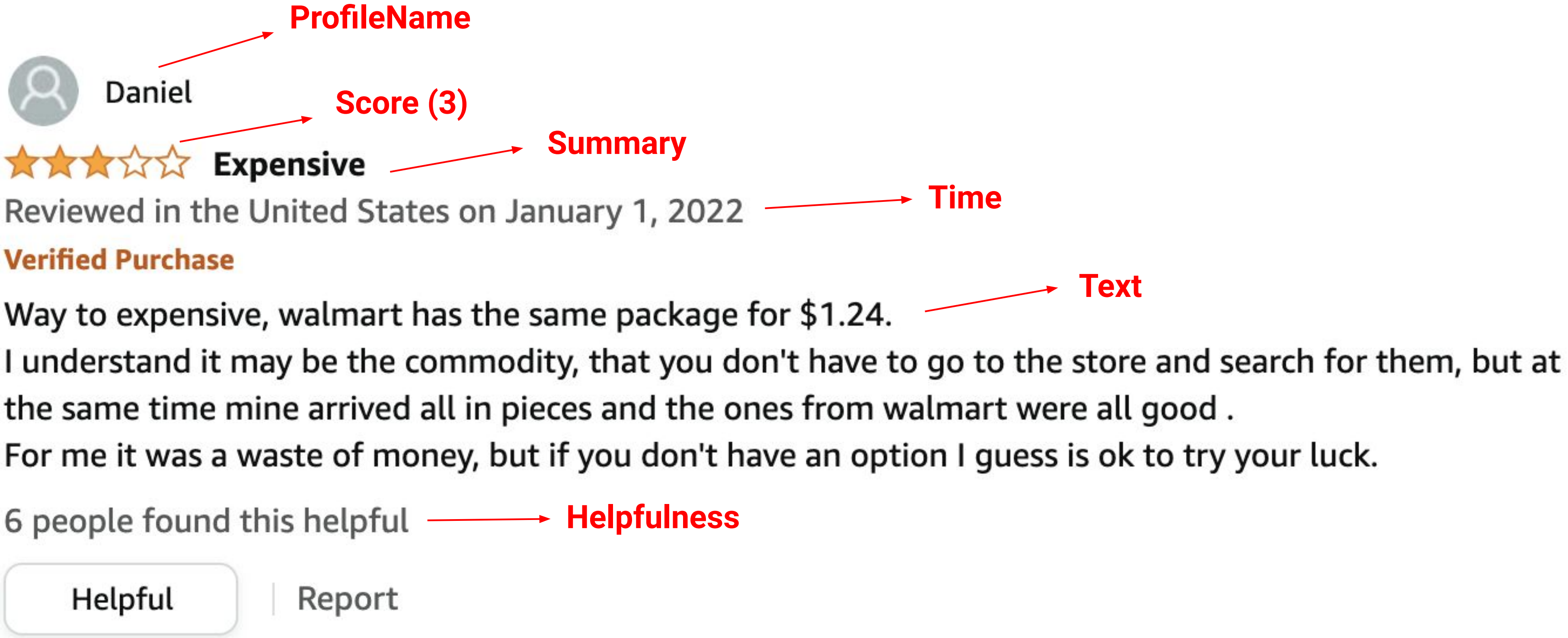
Example of dataset's structure and format

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...
...	...	...	...	...	...	...	...	...	...	...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	1299628800	Will not do without	Great for sesame chicken..this is a good if no...
568450	568451	B003S1WTCU	A3I8AFVPEE8KI5	R. Sawyer	0	0	2	1331251200	disappointed	I'm disappointed with the flavor. The chocolat...
568451	568452	B004I613EE	A121AA1GQV751Z	pksd "pk_007"	2	2	5	1329782400	Perfect for our maltipoo	These stars are small, so you can give 10-15 o...
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwel"	1	1	5	1331596800	Favorite Training and reward treat	These are the BEST treats for training and rew...
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfell17	0	0	5	1338422400	Great Honey	I am very satisfied ,product is as advertised,...

568454 rows × 10 columns

568454 rows x 10 columns

Example of an Amazon food product review



Dataset

- Utilized 568,454 Amazon food product reviews from Kaggle.
- Dataset includes 10 fields:
  `Id, ProductId, UserId, ProfileName, HelpfulnessNumerator, HelpfulnessDenominator, Score, Time, Summary, and Text.`

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...
...	...	...	...	...	...	...	...	...	...	...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	1299628800	Will not do without	Great for sesame chicken..this is a good if no...
568450	568451	B003S1WTCU	A3I8AFVPEE8KI5	R. Sawyer	0	0	2	1331251200	disappointed	I'm disappointed with the flavor. The chocolat...
568451	568452	B004I613EE	A121AA1GQV751Z	pksd "pk_007"	2	2	5	1329782400	Perfect for our maltipoo	These stars are small, so you can give 10-15 o...
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwel"	1	1	5	1331596800	Favorite Training and reward treat	These are the BEST treats for training and rew...
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfell17	0	0	5	1338422400	Great Honey	I am very satisfied ,product is as advertised,...

568454 rows × 10 columns

Data Preprocessing and Cleaning

- Removed duplicate reviews
- Focused on key columns:

  `Score` and `Text`

- Reduced to 393,933 unique data points

- Conducted distribution analysis

Data Preprocessing and Cleaning

- Adapted 5-category score system into a 3-class framework:

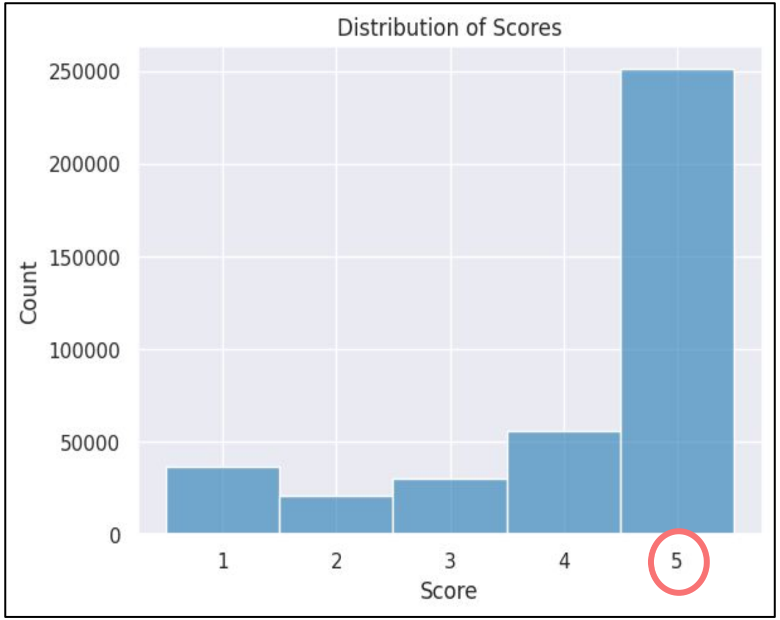
    Positive (scores > 3) → 1

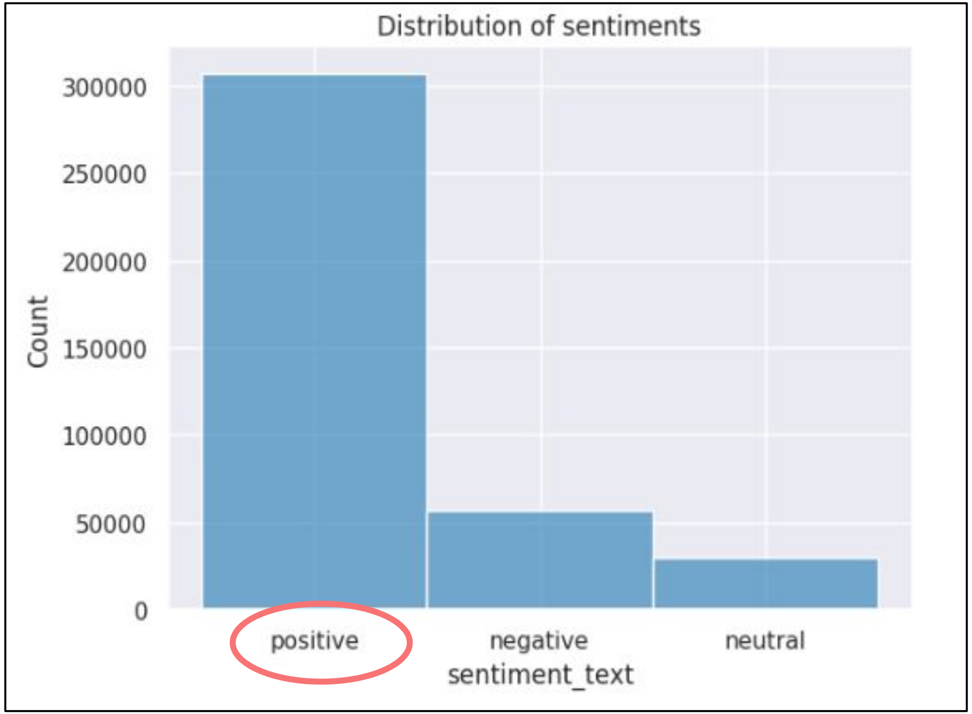
    Neutral (scores == 3) → 0

    Negative (scores < 3) → -1

- Reclassification aimed to address class imbalance.

Dataset's composition





Word Clouds for Each Sentiment Category

Positive

Negative

Neutral

Data Partitioning and Resampling Technique

- Dataset partitioning:

  The first **80%** for training and validation (80/20).

  The remaing **20%** reserved for model evaluation.

- Employed data undersampling for the majority class to address dataset imbalance

Feature Extraction Methods

- **Frequency-Based Techniques:**

➢ Bag of Words (BoW)
➢ TF-IDF

- **Word Embedding Techniques:**

➢ GloVe
➢ Word2Vec
➢ BERT Embeddings

Sentiment Classification Tool / Algorithms

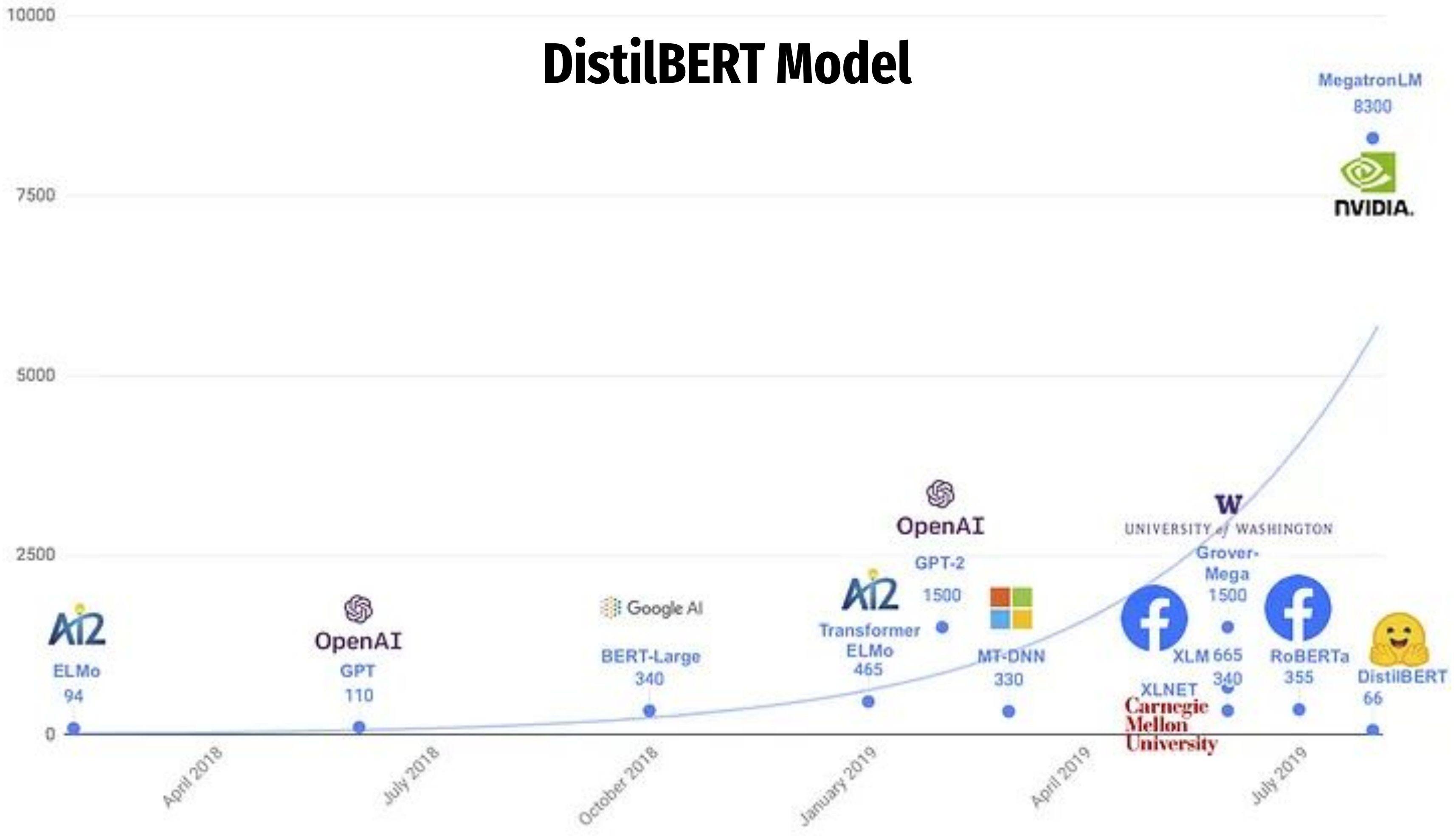
- **Sentiment Analysis Tool:**

➢ VADER (lexicon-based)

- **Machine Learning Classifiers:**

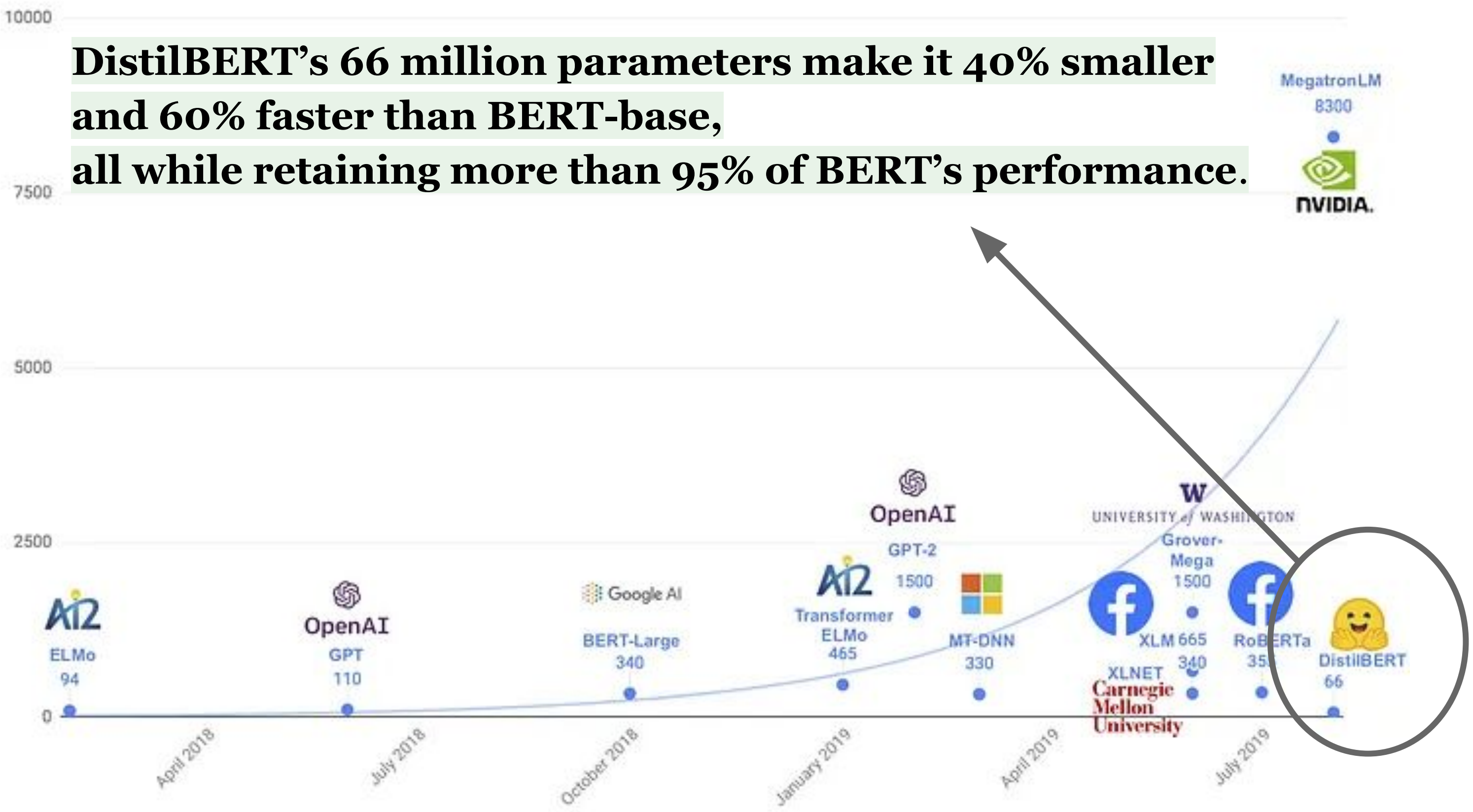
➢ Logistic Regression (LG)
➢ Naive Bayes (NB)
➢ Support Vector Machines (SVM)

DistilBERT Model



DistilBERT Model

DistilBERT's 66 million parameters make it 40% smaller and 60% faster than BERT-base, all while retaining more than 95% of BERT's performance.



Sanh et al., DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter (2019), arXiv:1910.01108

Finetuned DistilBERT Model

- **Hyperparameter Tuning in DistilBERT:**

➢ Manual tuning: learning rate, batch size, and training epochs
➢ 2 iterations due to computational limits

- **Regulation Strategies:**

➢ Weight decay to prevent overfitting
➢ Early Stopping to optimize training duration

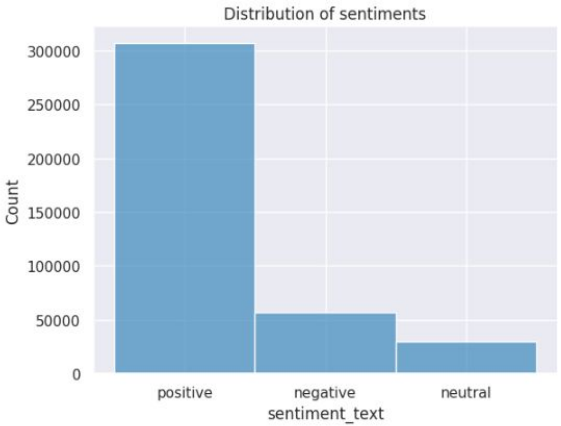
Model Evaluation Metrics

- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ F1-Score

Classification Report

Confusion Matrix

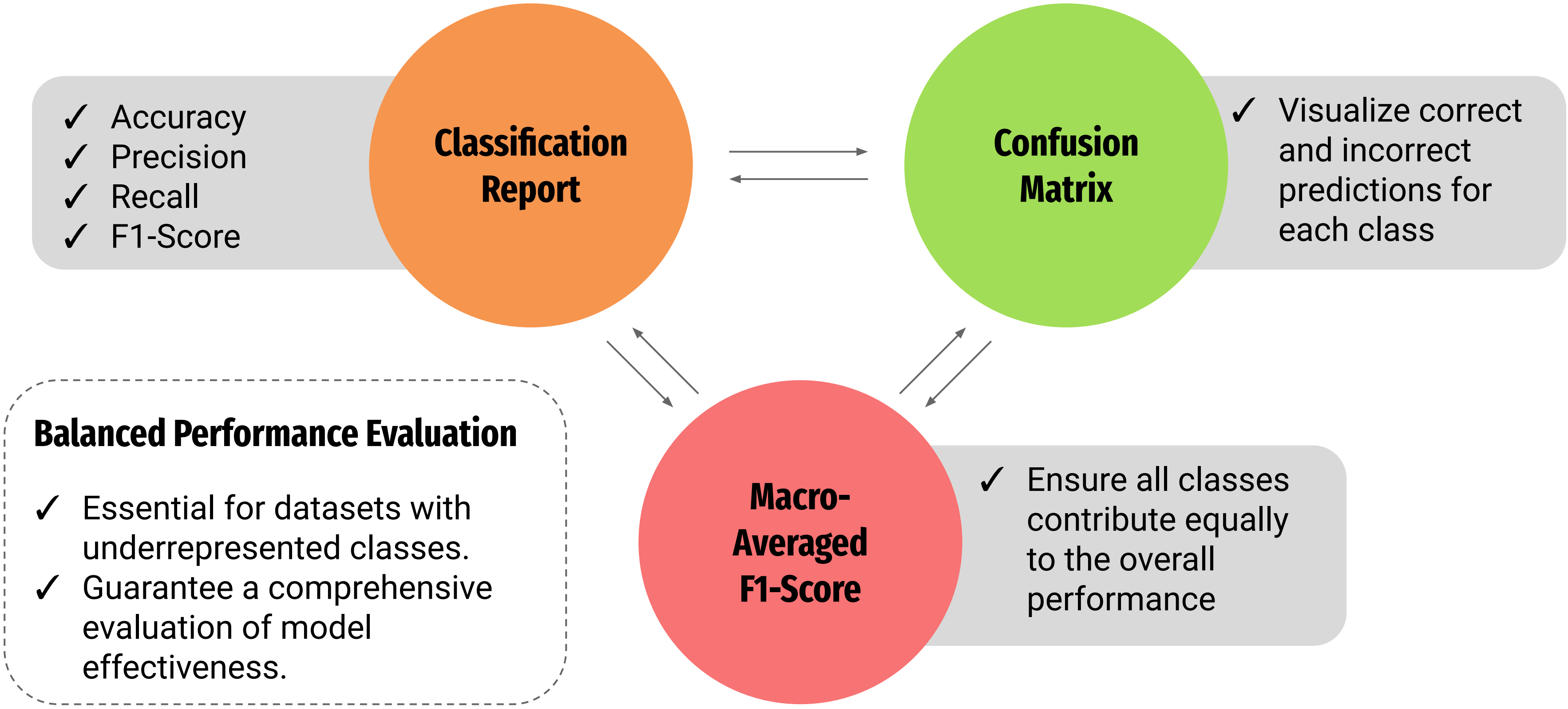
- ✓ Visualize correct and incorrect predictions for each class



Macro-Averaged F1-Score

- ✓ Ensure all classes contribute equally to the overall performance

Model Evaluation Metrics



Content

01 Introduction
- Background
- Overview
- Objectives

02 Related Works
- Previous studies
- Gaps in research

03 Methodology
- Data collection
- Feature extraction
- Classification algorithms
- DistilBERT fine-tuning
- Evaluation Metrics

04 Experiments
- Environment setup
- Process flow

05 Results
- Model performance
- Discussion analysis

06 Conclusion
- Ethics
- Limications
- Future work

Experiment Workflow

Setup

- ✓ Google Colab
- ✓ NVIDIA A100 GPU
- ✓ High-RAM
- ✓ Install libraries
- ✓ Data loading
- ✓ Data analysis
- ✓ Data preprocessing
- ✓ Data splitting
- ✓ Model Configs (Select hyperparameters)
- ✓ Monitoring (Track metrics, checkpoints)

Training & Validation

Training set: 80%
Validation set: 20%

Evaluation

- ✓ Macro-Averaged F1-Score
- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ Confusion Matrix

Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BoW + LG				
BoW + NB				
BoW + SVM				
BoW + LG + Undersampling				
BoW + NB + Undersampling				

Fig. 9

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
TF-IDF + LG				
TF-IDF + NB				
TF-IDF + SVM				
TF-IDF + LG + Undersampling				

Fig. 10

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GloVE + LG				
Word2Vec + LG				
BERT embedding + LG				

Fig. 11

Fig. 12

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VADER				
VADER + undersampling				

Fig. 13

Fine-tuned DistilBERT Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1st model				
2nd model				

Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BoW + LG	0.86	0.68	0.60	0.62
BoW + NB	0.83	0.62	0.61	0.62
BoW + SVM	0.86	0.70	0.57	0.58
BoW + LG + Undersampling	0.71	0.70	0.71	0.70
BoW + NB + Undersampling	0.70	0.71	0.70	0.70

Fig. 9

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
TF-IDF + LG	0.86	0.70	0.59	0.61
TF-IDF + NB	0.79	0.56	0.36	0.34
TF-IDF + SVM	0.83	0.71	0.46	0.48
TF-IDF + LG + Undersampling	0.68	0.68	0.68	0.68

Fig. 10

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GloVE + LG	0.82	0.60	0.48	0.49
Word2Vec + LG	0.82	0.60	0.48	0.49
BERT embedding + LG	0.84	0.70	0.52	0.55

Fig. 11

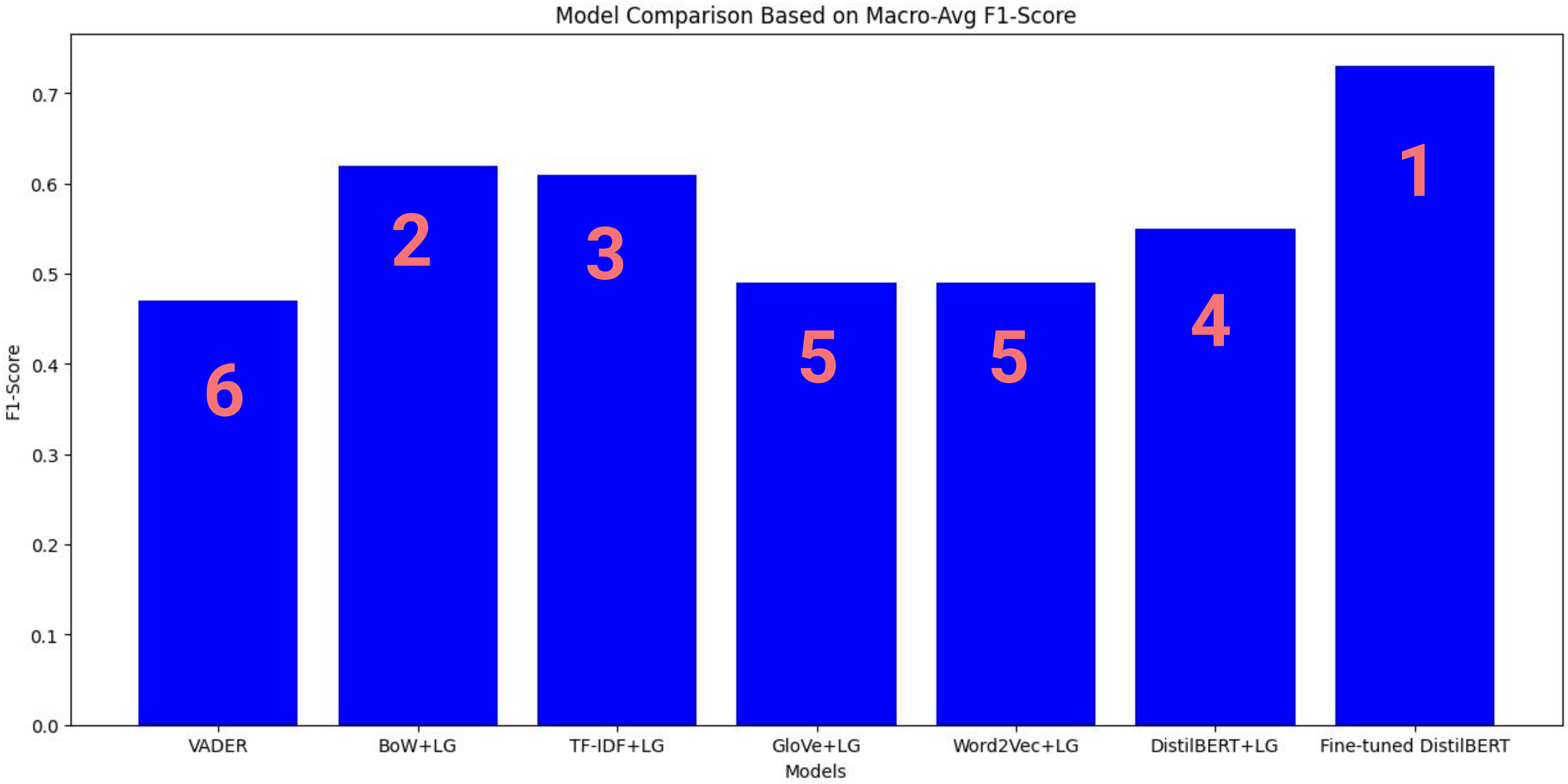
Fig. 12

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VADER	0.73	0.55	0.47	0.47
VADER + undersampling	0.58	0.54	0.46	0.46

Fig. 13

Fine-tuned DistilBERT Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1st model	0.78	0.26	0.33	0.29
2nd model	0.87	0.71	0.77	0.73

Model Performance



BoW outperforms TF-IDF 😏

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BoW + LG	0.86	0.68	0.60	0.62
BoW + NB	0.83	0.62	0.61	0.62
BoW + SVM	0.86	0.70	0.57	0.58
BoW + LG + Undersampling	0.71	0.70	0.71	0.70
BoW + NB + Undersampling	0.70	0.71	0.70	0.70

Fig. 9

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
TF-IDF + LG	0.86	0.70	0.59	0.61
TF-IDF + NB	0.79	0.56	0.36	0.34
TF-IDF + SVM	0.83	0.71	0.46	0.48
TF-IDF + LG + Undersampling	0.68	0.68	0.68	0.68

Fig. 10

- **BoW**

➢ Effective by capturing domain-specific terms

➢ Effective by focusing on word presence (not context)

- **TF-IDF**

➢ Less effective from contextual weighting due to review lengths and specific terms in food product reviews

BoW outperforms Word Embeddings 😱

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BoW + LG	0.86	0.68	0.60	0.62
BoW + NB	0.83	0.62	0.61	0.62
BoW + SVM	0.86	0.70	0.57	0.58
BoW + LG + Undersampling	0.71	0.70	0.71	0.70
BoW + NB + Undersampling	0.70	0.71	0.70	0.70

Fig. 9

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GloVE + LG	0.82	0.60	0.48	0.49
Word2Vec + LG	0.82	0.60	0.48	0.49
BERT embedding + LG	0.84	0.70	0.52	0.55

Fig. 11

- **BoW**
  - ➢ Effective by aligning dataset's straightforward characteristic
  - ➢ Effective by handling negations as distinct features
- **Word Embeddings**
  - ➢ Less effective for specific sentiment contexts in food product reviews
  - ➢ Less effective due to potential classifier confusion

BoW outperforms VADER 😄

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BoW + LG	0.86	0.68	0.60	0.62
BoW + NB	0.83	0.62	0.61	0.62
BoW + SVM	0.86	0.70	0.57	0.58
BoW + LG + Undersampling	0.71	0.70	0.71	0.70
BoW + NB + Undersampling	0.70	0.71	0.70	0.70

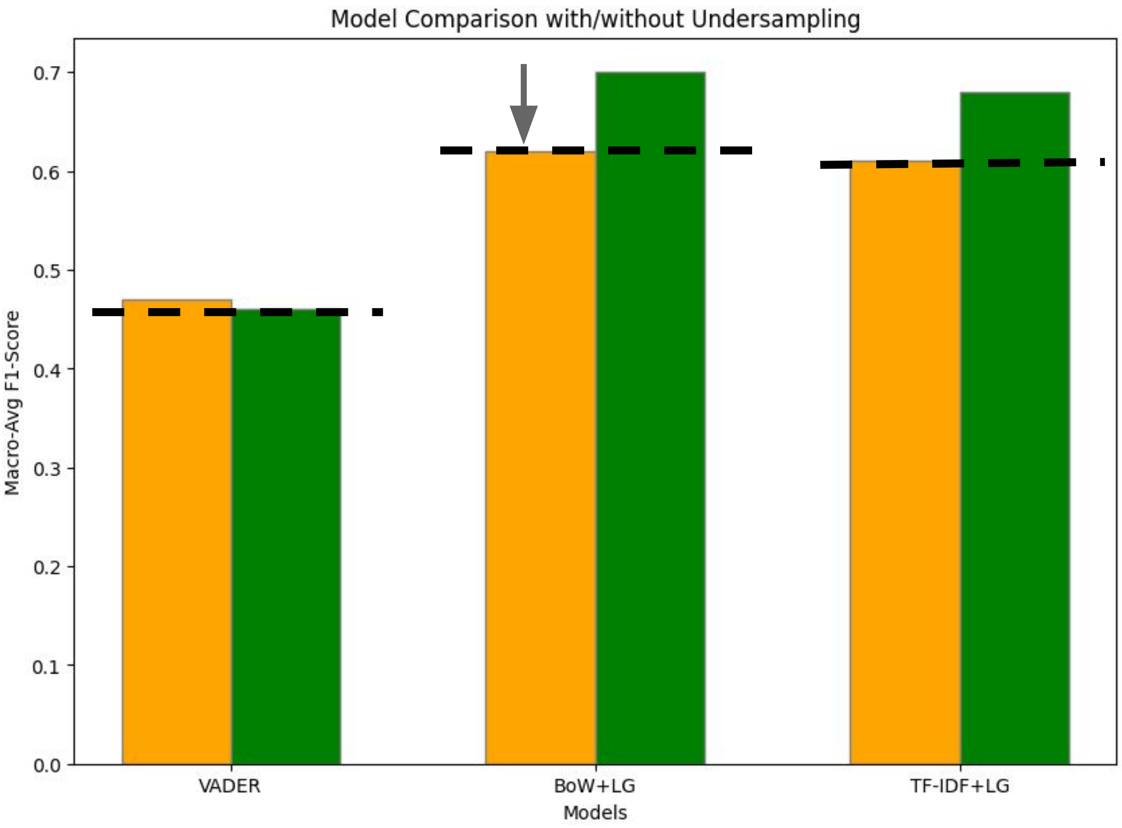
Fig. 9

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VADER	0.73	0.55	0.47	0.47
VADER + undersampling	0.58	0.54	0.46	0.46

Fig. 12

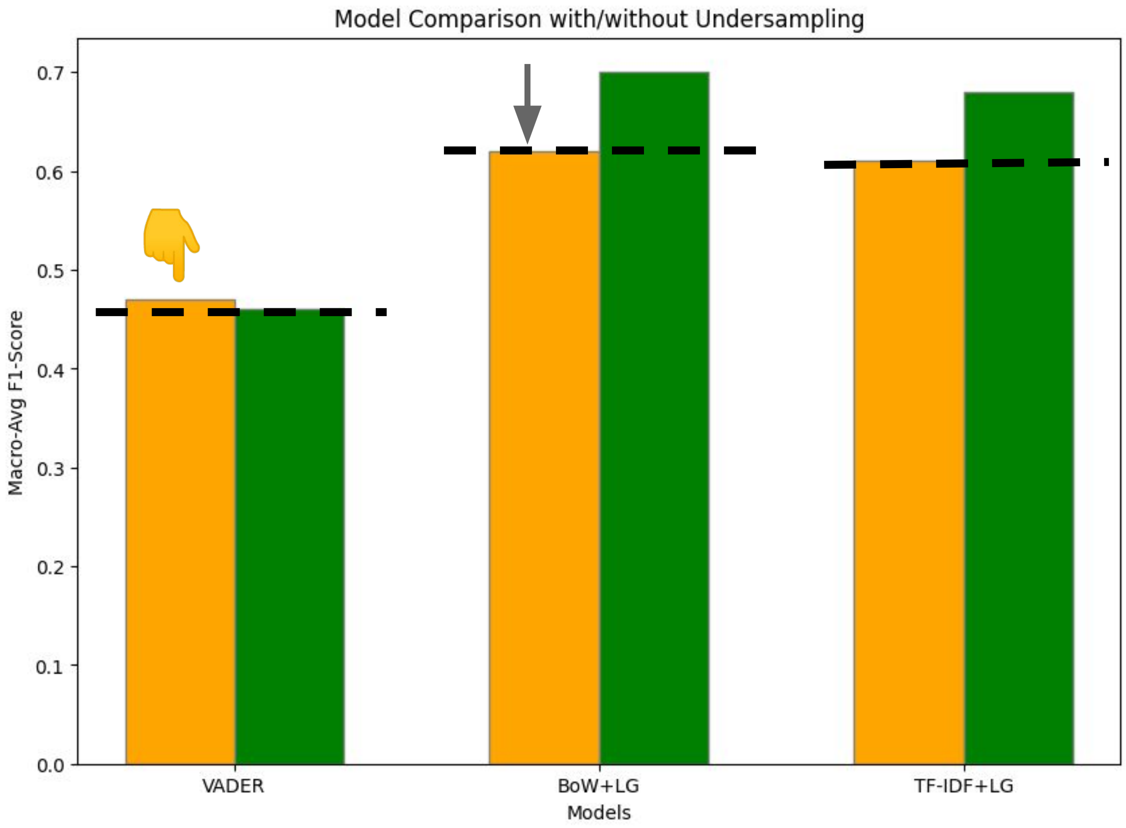
- **BoW + LG**
  - ➢ Effective after training on the specific dataset
  - ➢ Effective in understanding unique nuances and vocab
- **VADER**
  - ➢ Less effective due to predefined rules and scores for sentiment intensity

Impact of Undersampling Strategy 👍



- **Mitigate Data Imbalance**
- **Enhance Model Efficacy**

Impact of Undersampling Strategy 👍



- **Mitigate Data Imbalance**
- **Enhance Model Efficacy**

Implication:

Consider the specific operational mechanisms when applying dataset manipulation techniques.

Model Performance

DistilBERT Model Fine-tuning

Parameters	1st training	2nd training
Learning rate	2e-4	5e-5
Batch size	8	8
Number of epochs	4	5
Weight decay	0.01	0.01
Seed	0	0
Optimizer	AdamW	AdamW

- **Lower learning rate**

➢ More precise weight adjustments

- **More training epochs**

➢ Deeper learning from the data patterns

DistilBERT Model Fine-tuning

1st iteration

```
trainer.train()
```

[7500/23480 1:52:36 < 3:59:59, 1.11 it/s, Epoch 1/4]

Step	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
500	1.008900	0.756776	0.784388	0.261463	0.333333	0.293056
1000	1.010300	0.772518	0.784388	0.261463	0.333333	0.293056
1500	1.014700	0.789708	0.784388	0.261463	0.333333	0.293056
2000	1.013800	0.797439	0.784388	0.261463	0.333333	0.293056
2500	1.011200	0.750883	0.784388	0.261463	0.333333	0.293056
3000	1.020800	0.780068	0.784388	0.261463	0.333333	0.293056
3500	1.019000	0.776988	0.784388	0.261463	0.333333	0.293056
4000	1.017200	0.789884	0.784388	0.261463	0.333333	0.293056
4500	1.005400	0.743515	0.784388	0.261463	0.333333	0.293056
5000	1.012300	0.733742	0.784388	0.261463	0.333333	0.293056
5500	1.006400	0.808375	0.784388	0.261463	0.333333	0.293056
6000	1.005700	0.746866	0.784388	0.261463	0.333333	0.293056
6500	1.010000	0.785966	0.784388	0.261463	0.333333	0.293056
7000	1.015400	0.761455	0.784388	0.261463	0.333333	0.293056
7500	1.013800	0.764767	0.784388	0.261463	0.333333	0.293056

- **Lower learning rate**
  - ➢ More precise weight adjustments
- **More training epochs**
  - ➢ Deeper learning from the data patterns

DistilBERT Model Fine-tuning

Parameters	1st training	2nd training
Learning rate	2e-4	5e-5
Batch size	8	8
Number of epochs	4	5
Weight decay	0.01	0.01
Seed	0	0
Optimizer	AdamW	AdamW

- **Lower learning rate**

➢ More precise weight adjustments

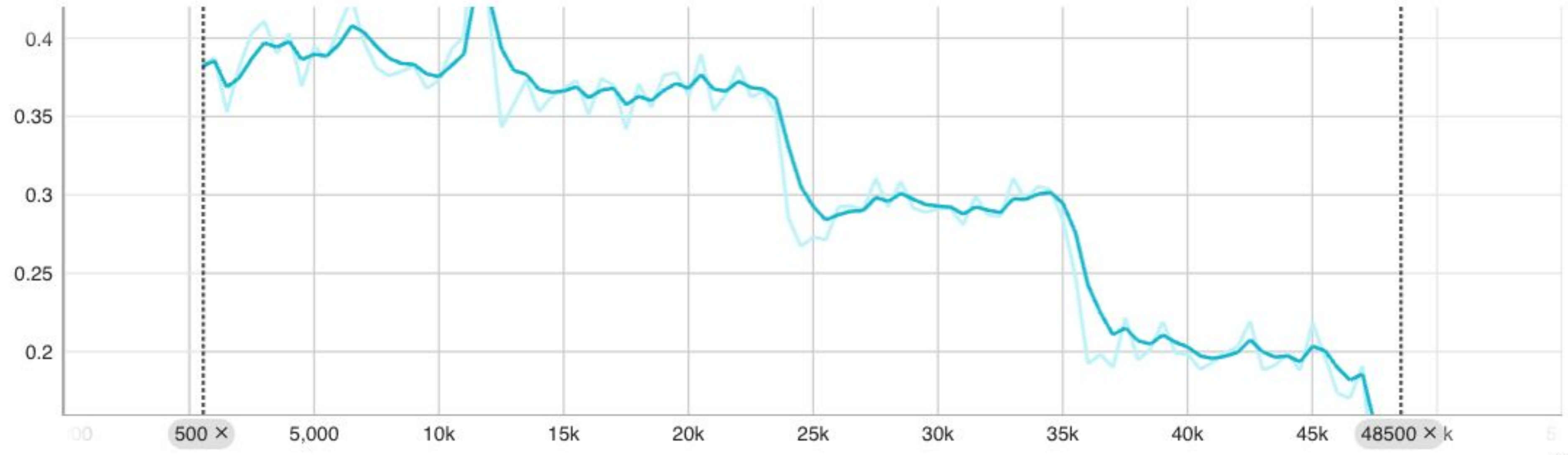
- **More training epochs**

➢ Deeper learning from the data patterns

DistilBERT Model Fine-tuning

2nd iteration

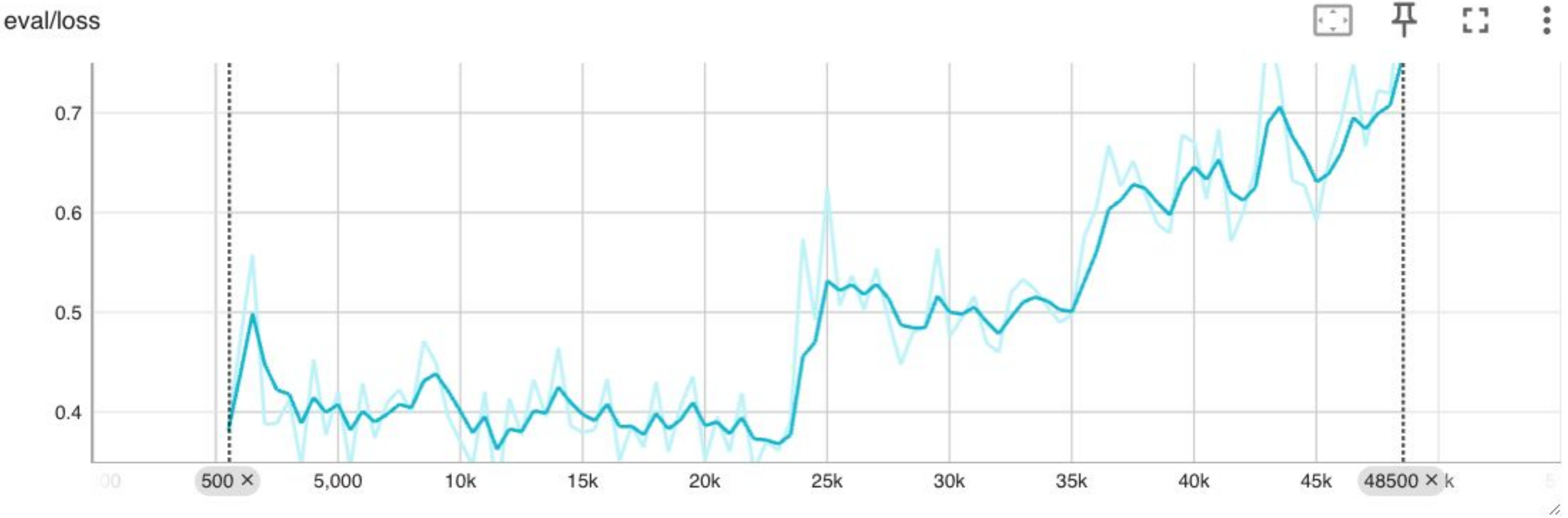
train/loss



Run ↑	Min	Max	Start Value	End Value	△Value	△%	Start Step	End Step
runs/Nov28_17-56-46_70b26b9aa4af	0.1178	0.4334	0.3819	0.1178	↓0.2641	↓-69%	500	48,500

DistilBERT Model Fine-tuning

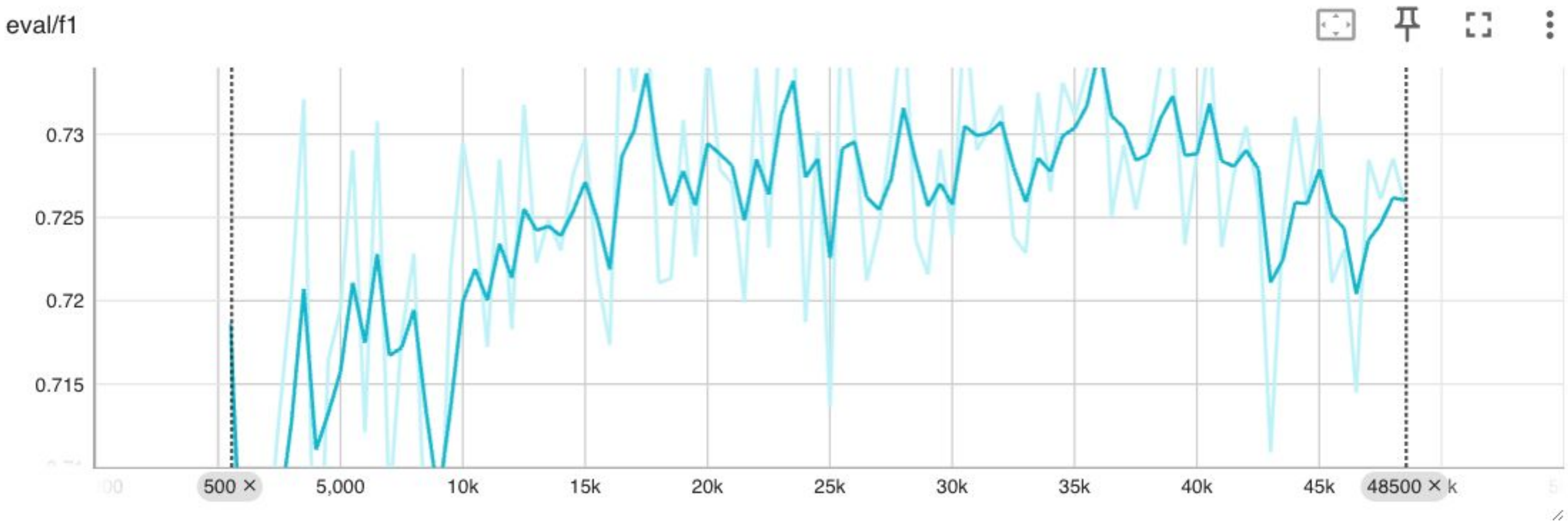
2nd iteration



Run ↑	Min	Max	Start Value	End Value	△Value	△%	Start Step	End Step
runs/Nov28_17-56-46_70b26b9aa4af	0.3623	0.7539	0.3804	0.7539	↑0.3734	↑98%	500	48,500

DistilBERT Model Fine-tuning

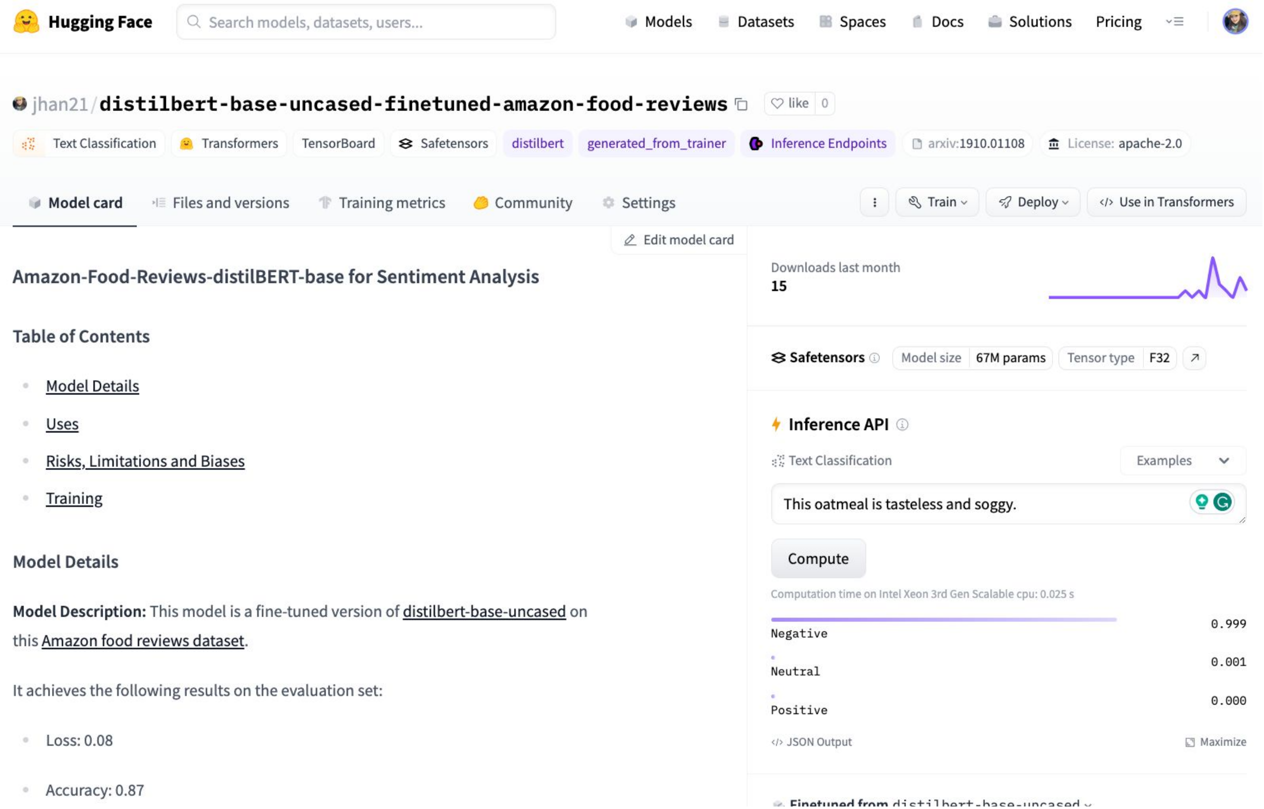
2nd iteration



Run ↑	Min	Max	Start Value	End Value	∆Value	∆%	Start Step	End Step
runs/Nov28_17-56-46_70b26b9aa4af	0.7003	0.7351	0.7186	0.726	↑0.0074	↑1%	500	48,500

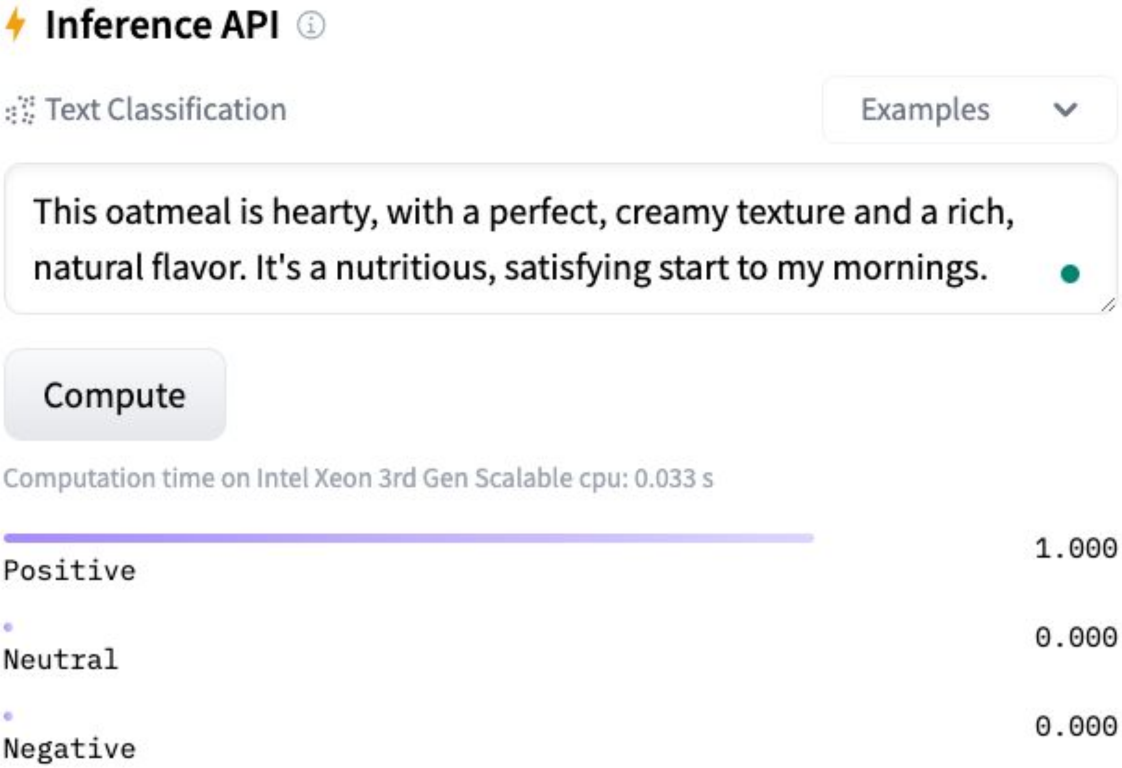
Fine-tuned DistilBERT Model Card

https://huggingface.co/jhan21/distilbert-base-uncased-finetuned-amazon-food-reviews



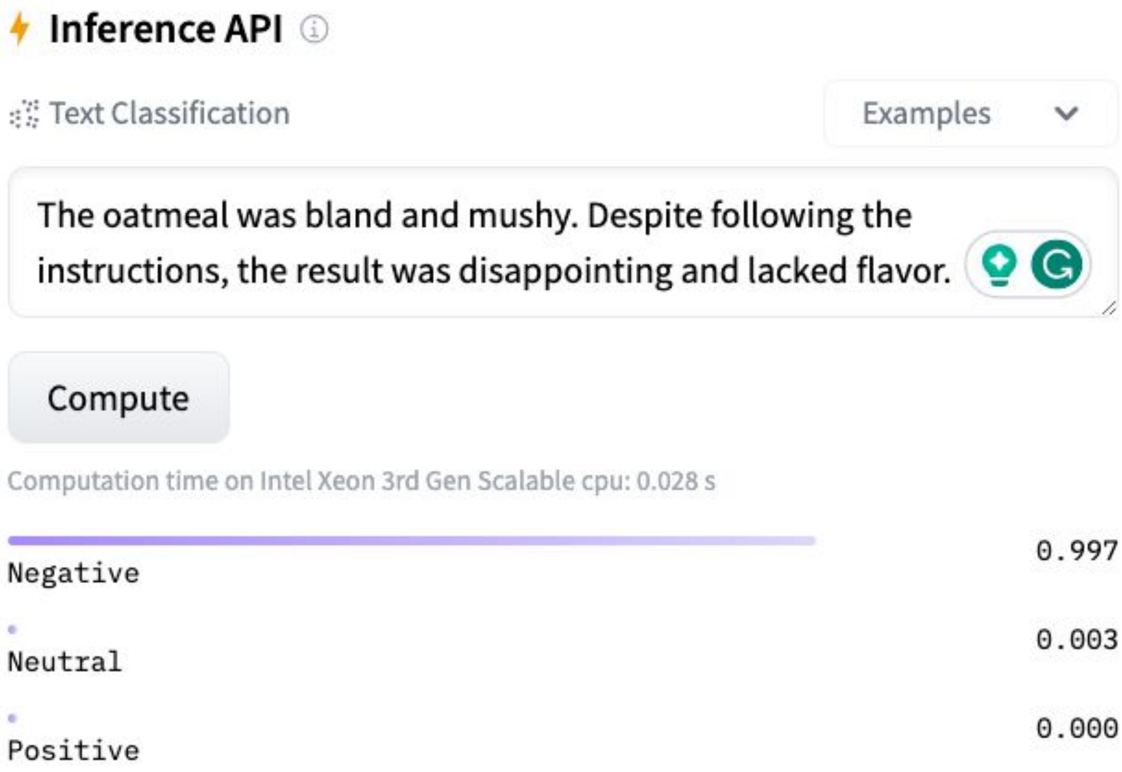
Sample outputs

https://huggingface.co/jhan21/distilbert-base-uncased-finetuned-amazon-food-reviews



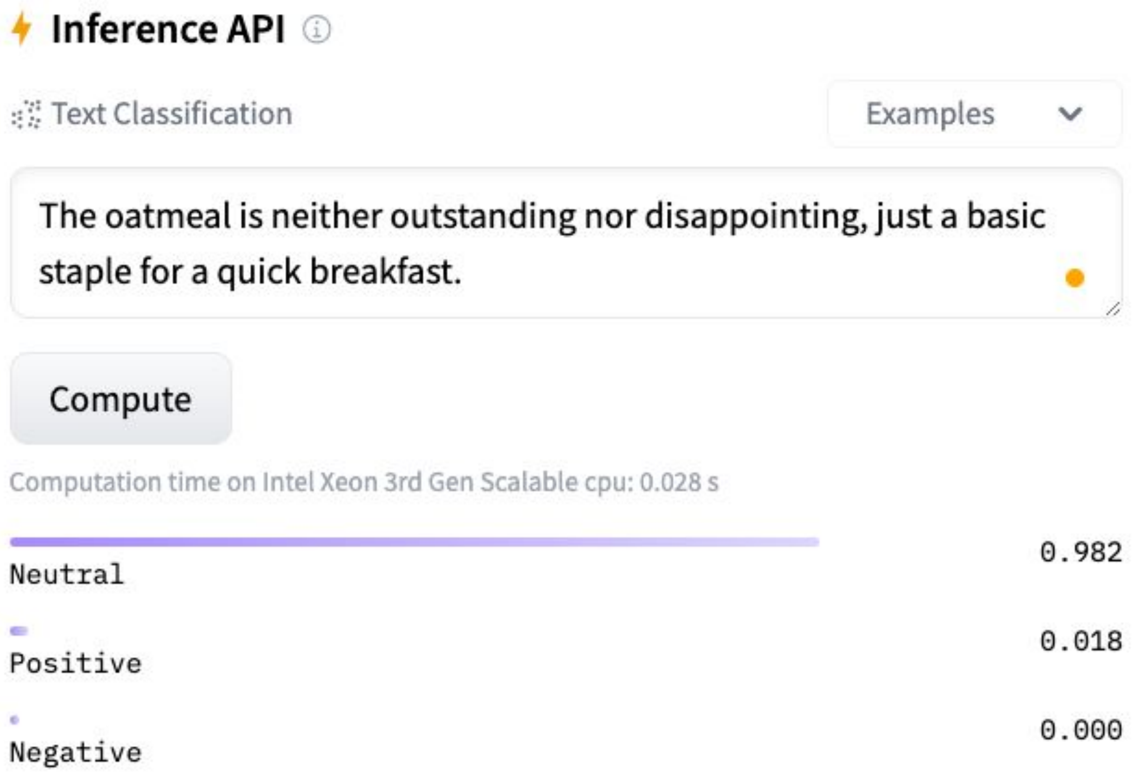
Sample outputs

https://huggingface.co/jhan21/distilbert-base-uncased-finetuned-amazon-food-reviews



Sample outputs

https://huggingface.co/jhan21/distilbert-base-uncased-finetuned-amazon-food-reviews



Sample outputs

https://huggingface.co/jhan21/distilbert-base-uncased-finetuned-amazon-food-reviews

Inference API

Text Classification | Examples

This oatmeal is okay, but I will not buy it again.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.024 s

Neutral	0.995
Negative	0.004
Positive	0.001

Inference API

Text Classification | Examples

This oatmeal is tasteless and soggy.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.025 s

Negative	0.999
Neutral	0.001
Positive	0.000

Conclusion

- Problem Addressed
- Simplicity Over Complexity
- Importance of Hyperparameter Optimization
- Strategic Insights and practical implications

Ethics

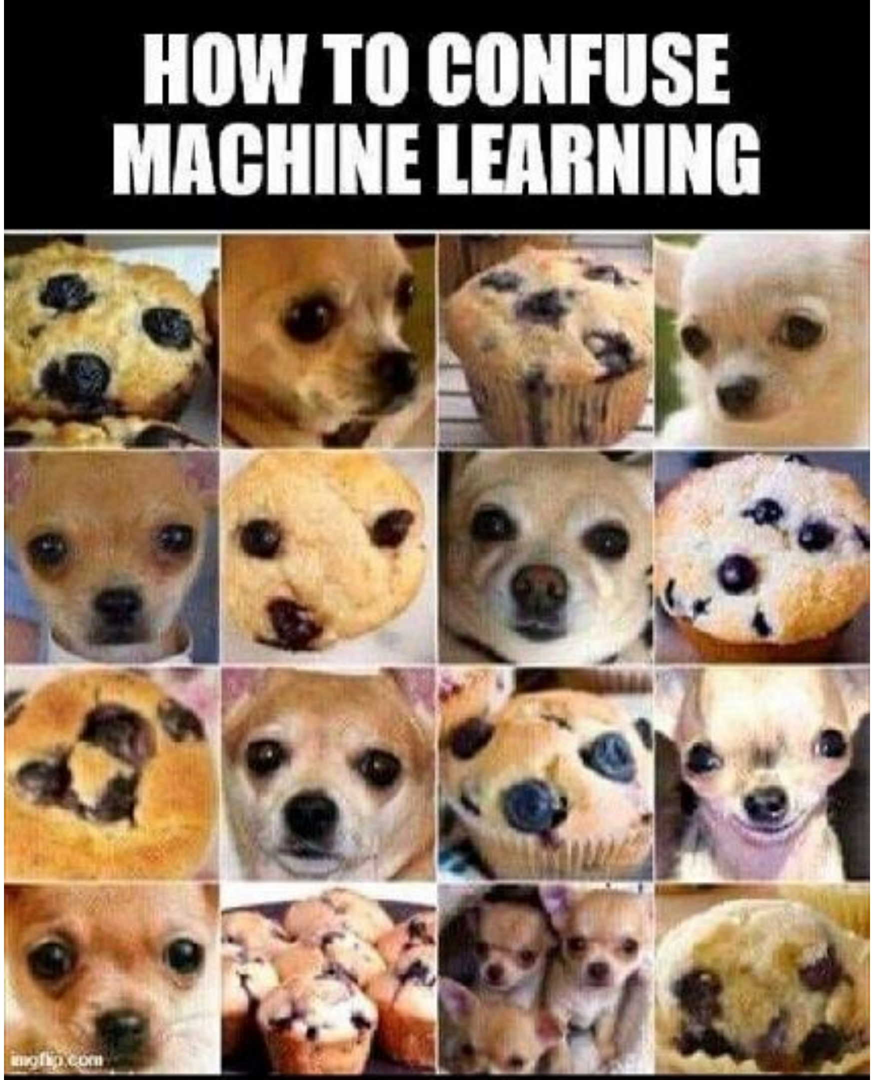
- Data Privacy Compliance
- Transparency and Reproducibility
- Social Impact and Ethical Use
- Future Ethical Considerations

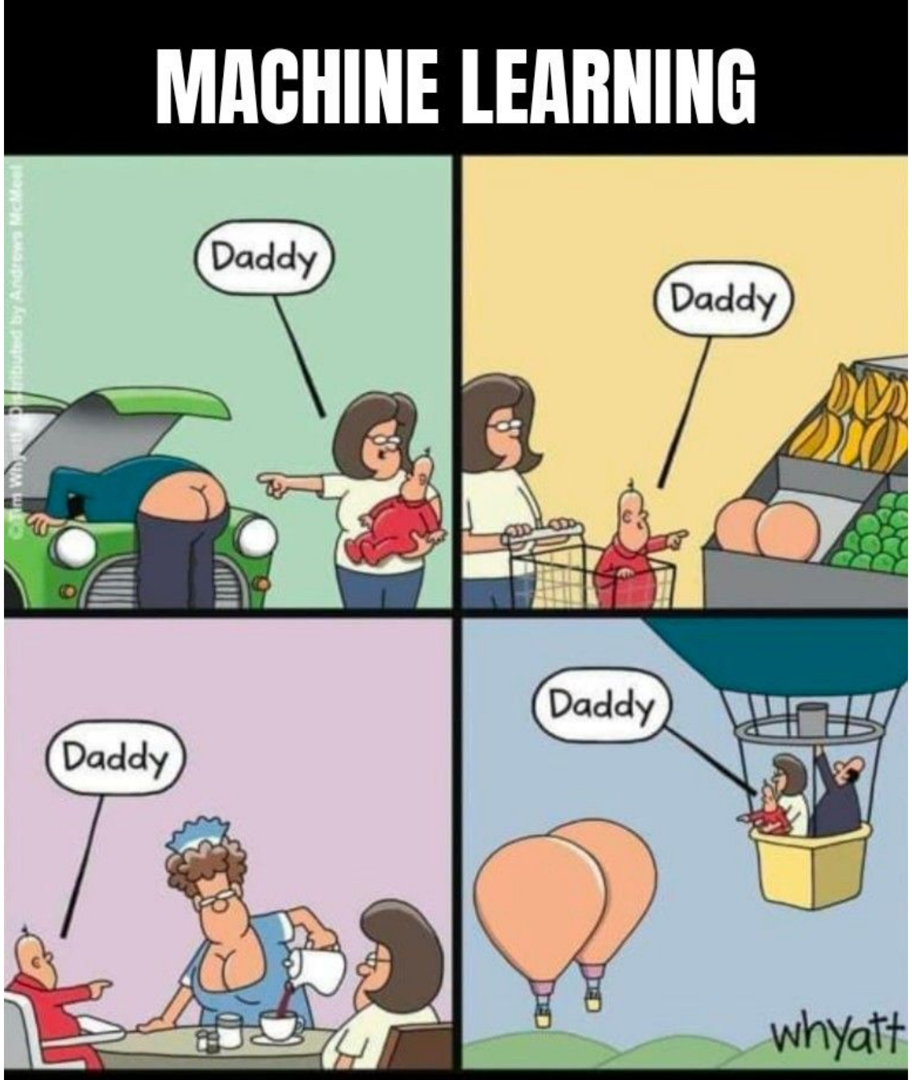
Limitations & Future Work

- Computational Constraints
- Dataset Specificity Limitations
- Scope of Methodologies
- Data Manipulation and Hyperparameter
- Future work to improve model performance:
  cross-validation, regulation, random search/grid search

Fine-tuning a pre-trained Model is like "voodoo".
— Dr. Church

- Lack of predictability
- Intuition and experience
- Sensitivity to small changes
- Trials and error process


HOW TO CONFUSE
MACHINE LEARNING
imgflip.com


MACHINE LEARNING
© Tim Whyatt Distributed by Andrews McMeel
Daddy
Daddy
Daddy
Daddy
whyatt

References

Yaylı, A., & Bayram, M. (2012). E-WOM: The effects of online consumer reviews on purchasing decisions. *International Journal of Internet Marketing and Advertising*, 7(1), 51-64.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(1), 1-14.

Anees et al. (2020). Survey paper on sentiment analysis: Techniques and challenges. *EasyChair*, 2516-2314.

Xu et al. (2015). Sentiment analysis of yelp's ratings based on text reviews. In *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (Vol. 17, No. 1, pp. 117-120).

Sharma, A., & Dey, S. (2012). A comparative study of feature selection and machine learning techniques for sentiment analysis. In *Proceedings of the 2012 ACM research in applied computation symposium* (pp. 1-7).

O'Keefe, T., & Koprinska, I. (2009). Feature selection and weighting methods in sentiment analysis. In *Proceedings of the 14th Australasian document computing symposium, Sydney* (pp. 67-74).

Shrestha, N., & Nasoz, F. (2019). Deep learning sentiment analysis of amazon. com reviews and ratings.

Elmurngi, E. I., & Gherbi, A. (2018). Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques. *J. Comput. Sci.*, *14*(5), 714-726.

Yang et al. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, *8*, 23522-23530.

Agarwal at al.. (2011). Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media*.

Pandarachalil et al. (2015). Twitter sentiment analysis for large-scale data: an unsupervised approach. *Cognitive computation*, 254-262.

Fujino et al. (2008). Multi-label text categorization with model combination based on f1-score maximization. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*.

Goldberg, Y., & Hirst, G. (2017). Neural network methods in natural language processing. morgan & claypool publishers.

Shelke et al. (2017). A review on imbalanced data handling using undersampling and oversampling technique. *Int. J. Recent Trends Eng. Res*, *3*(4), 444-449.

Drummond, C., & Holte, R. C. (2003). C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on learning from imbalanced datasets II* (Vol. 11, pp. 1-8).

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 122, p. 16).

Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, *181*(1), 25-29.

Sanh et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Li et al. (2020). Rethinking the hyperparameters for fine-tuning. *arXiv preprint arXiv:2002.11770*.